DESIGNING MACHINE LEARNING TECHNIQUE FOR REDUCING CATASTROPHIC FORGETTING IN CONTINUAL LEARNING

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Manav Rachna International Institute of Research and Studies

(Deemed to be University)

by

Neeraj

Registration number: 22-23/Ph.D/W/018

Under the supervision of

Dr. Poonam Nandal

Professor

Department of Computer Science and Engineering

School of Engineering & Technology

Manav Rachna International Institute of Research & Studies

Sector-43, Surajkund–Delhi Road, Faridabad– 121004

SCHOOL OF ENGINEERING AND TECHNOLOGY

Certificate

I hereby declare that the Short Synopsis titled "Designing Machine Learning Technique for Reducing Catastrophic Forgetting in Continual Learning" being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the Department of Computer Science and Engineering under the School of Engineering & Technology of Manav Rachna International Institute of Research & Studies, Faridabad during the academic year 2024-25. It is a bonafide record of original work carried out under the supervision of Dr. Poonam Nandal and has not been submitted in full or part to any other university or institute for a degree or diploma. Further, it is certified that Short Synopsis is plagiarism-free and proof has been submitted.

Neeraj

22-23/Ph.D/W/018

11/11/25

Dr. Poonam Nandal

Professor, CSE

School of Engineering & Technology

(Chairperson, DRC) 11.11.25

(Dean, School of Engineering & Technology)

Abstract

Artificial intelligence is a field of computer science that depends on the ability of a machine to learn tasks sequentially, which is called continual learning. Continual learning, also known as incremental learning, where the systems are supposed to learn new knowledge progressively over time without forgetting earlier learned knowledge. One of the main obstacle to continual learning is catastrophic forgetting, in which the neural network model, when learning various tasks sequentially, loses its previously learned knowledge from dynamic data distributions and degrades the performance of earlier learned tasks. This problem has widely occurred in the connectionist model. The neural network training and stability-plasticity dilemma are affected by catastrophic forgetting, where the stability-plasticity dilemma states that when the network becomes too plastic, it lacks the stability required to prevent the previous knowledge from being forgotten. Further, it is vital to reduce catastrophic forgetting in environments where the model must perform in dynamic, real-world situations, guaranteeing that machine learning systems maintain consistent, reliable performance across multiple tasks. There are several existing methods such as replay, regularization, and modular network architecture, to reduce catastrophic forgetting. Still, none of them has been able to successfully and completely reduce the issue in all tasks and domains. This research aims to examine the existing methods of reducing catastrophic forgetting, as well as their strengths and weaknesses, and propose a hybrid technique that enhances the retention of information while learning new tasks. The proposed hybrid technique would integrate existing methods in a balanced way to address their weaknesses effectively and offer flexibility. The proposed technique would use a hybrid model, evaluated on benchmark datasets, and further will be compared with other state of the art techniques.

Keywords: Artificial Intelligence, Machine Learning, Catastrophic Forgetting, Continual Learning, Deep Neural Network.

Table of Contents

	Certificate	i
	Abstract	ii
	Table of Contents	iii-iv
S. No.	**Contents**	**Page No.**
1	Description of Broad Area	1-8
	1.1 Overview of Machine Learning	1
	1.1.1 Machine Learning Methods	
	1.2 Continual Learning	4
	1.3 Catastrophic Forgetting	5
	1.3.1 Factors of Catastrophic Forgetting	
	1.4 Addressing Forgetting	6
	1.4.1 Task-Aware	
	1.4.2 Task-Free	
	1.4.3 Online and Offline	
	1.4.4 Semi-Supervised, Few-shot and Unsupervised	
	1.5 Motivation	8
2	Literature Review	8-14
3	Research Gaps	15
4.	Problem Statement	15
5	Research Objectives	16
6	Research Methodology	16
7	Expected Outcome of Research	16

References 17

Similarity Verification Report

Grammarly Certificate

1. DESCRIPTION OF BROAD AREA

The emergence of Artificial Intelligence (AI) and Machine Learning (ML) is due to the limitations in traditional systems like programmatic rigidity, manual feature engineering, complex decision-making, lack of automation, and poor performance in complex tasks, etc. These limitations have led to the development of more intelligent and adaptive approaches. AI is more often employed to create intelligent agents that can carry out different jobs. Machine Learning emphasizes creating algorithms that let computers learn from data and make decisions without explicit programming [1]. Machine learning enhances the automation of several tasks that are not done by traditional systems, but it faces some problems in handling complex and vast amounts of data. Further, the feature engineering in ML is done manually. The researchers have used deep learning in which feature engineering is done automatically. Deep learning uses multiple-layer neural networks to learn hierarchical representations of input data. It automatically identifies abstract and complicated features from raw data without the need for explicit feature engineering. In the era of intelligent systems, continual learning is a key for neural network models that constantly learn from new data streams. It maintains model performance in a dynamic environment, adapting to changing data distributions, and efficiently utilizing computational resources. It also makes neural network models more versatile and applicable in a wide range of real-world scenarios. The field of continual learning within machine learning offers a distinct set of challenges as well as remarkable adaptability for intelligent systems. In dynamic contexts where models need to evolve and adapt to new data, continual learning is essential. One major obstacle to achieving the full potential of continual learning paradigms is forgetting. This research focuses on the challenges of retaining past experiences that arise in continual learning while adapting to new information.

1.1 OVERVIEW OF MACHINE LEARNING

Machine learning (ML) is a subset of Artificial Intelligence (AI) where the computer automatically learns without being explicitly programmed. Its main function is to develop algorithms and models that enable a machine to learn from data and provide better results according to past observations or data records.

ML performs various complex tasks easily by accessing huge amounts of data [1]. It is used in various application areas like image and speech recognition, traffic prediction, self-driving cars, natural language processing, recommendation systems, fraud detection, email spam filtering, stock market trading, automated tasks, etc.

1.1.1 MACHINE LEARNING METHODS

In this subsection, various machine learning methods are discussed below:

i) **Supervised Learning:** In this learning method, the machine is supervised or trained by using a "labeled" dataset. The machine forecasts the output based on training [1]. There are two types of supervised learning, classification, and regression as discussed below:

Classification: In this method, data is distributed into predefined classes or labels present in the dataset. It is valid where the output variable is categorical. The real-world applications of classification algorithms are Spam Detection and Email filtering. Some of the classification algorithms are Support Vector Machine (SVM), Decision Tree, Logistic Regression, K-Nearest Neighbours (KNN), Random Forest and Naive Bayes, etc.

Regression: It predicts continuous numerical output based on input features such as weather prediction. Here the linear relationship exists between input and output variables. Regression algorithms include Linear, Polynomial, Ridge, Lasso, Decision tree algorithm, and Random Forest regression etc.

These learning models have high accuracy because of training with labeled data. It is used in real-world applications like Image classification, Natural Language Processing (NLP), speech recognition, predictive analytics, fraud detection, email spam detection, quality control in manufacturing, gaming and weather forecasting, etc.

ii) **Unsupervised Learning:** In this learning method, the unlabelled dataset is used to train the model, and output is predicted by the machine without any supervision based on past experiences and learns from the previously introduced features [1]. It is used to find hidden patterns and clusters that exist in given data; it is mainly used in data exploration, visualization, dimensionality reduction, etc. There are two types of unsupervised learning clustering and association as discussed below:

Clustering: It is the process of making clusters or groups based on similar features of data. It identifies the patterns and relationships without any requirement of labeled examples. It includes various algorithms like K-Means Clustering, Principal Component Analysis (PCA), Independent Component Analysis and Mean shift, etc.

Association: It is used to find the relationship between variables in the dataset. The real-world application area is Market Basket analysis and Web usage mining etc. The algorithms like Apriori, Eclat, and FP-growth are the most widely used.

Unsupervised Learning helps to identify relationships and hidden patterns between the data. It is used in Customer Segmentation, Anomaly Detection, Data Exploration, Dimensionality Reduction, Recommendation Systems, Topic Modelling, Density Estimation, Image and Video Compression, Data Preprocessing and Market Basket analysis, etc.

iii) **Semi-supervised Machine Learning:** It is used when both supervised and unsupervised learning is inadequate due to the presence of less amount of labeled data than unlabeled data. In such a situation, the inadequate unlabeled data is used to deduce information about them. This method is called semi-supervised learning [1]. It includes co-training, self-training, label propagation, Generative adversarial networks (GANs), and Graph-based semi-supervised learning as discussed below:

Co-training: In this technique, two different models are trained on different subsets of the unlabeled data, and then these models label one another's predictions.

Self-training: In this technique, the model training is done on labeled data and then the prediction is performed with unlabeled data.

Label propagation: In this technique, labels are provided to unlabeled data by the propagation of labels from a small set of known labeled data points to a large set of unknown unlabeled data points.

Generative adversarial networks (GANs): This technique generates artificial data by adversarial training of two neural networks, a generator and a discriminator. The generated data is used for unsupervised or semi-supervised learning.

Graph-based semi-supervised learning: It represents the relationships between the data points. The propagation is from the labeled to the unlabeled data points.

Semi-supervised learning produces better generalization, but it is more complex. In this learning small of labelled data is required. It is used in various application areas like Image Classification, Object Recognition, Natural Language Processing (NLP), Speech Recognition, Recommendation Systems, Healthcare and Medical Imaging.

iv) **Reinforcement Machine Learning:** In this learning, learning agents learn through a reward-based approach. The agent aims to follow the shortest and most accurate path to achieve the goal, though there are start and finish points [1]. When agent follows the proper procedures they get positive rewards, but if they follow the wrong path, they get negative rewards.

Application areas of Reinforcement learning are Game Playing (AlphaGO and AlphaGO Zero), Resource Management, Robotics, Text Mining, Adaptive Personal Assistants, Autonomous

Vehicles, Recommendation Systems, Healthcare, Energy Management, Virtual and Augmented Reality (AR). This learning is classified into two types Positive Reinforcement and Negative reinforcement learning. It includes various learning algorithms like Q-Learning, Deep Q- Learning, and SARSA(State-Action-Reward-State-Action).

Deep Learning which is a more specified field of machine learning uses all the above methods of machine learning to train models on sequential data. Continual Learning allows the model to update when new data is given whereas traditional models would become outdated in the same situation.

1.2 CONTINUAL LEARNING

Continual Learning (CL) [2] is called Life-long Learning, incremental Learning, or sequential Learning. In this learning new tasks are learned sequentially. CL can be classified based on task information: no. of processing passes and amount of labeled data available. Task information includes explicit task definitions such as task labels or IDs. The CL based on task information are task-aware and task-free. In task-aware learning, explicit information about the tasks is available during training, which helps address forgetting. The availability of task information is helpful in the development of various techniques to mitigate forgetting. In [2] CL is classified into task incremental, domain incremental, and class incremental learning. In task-free learning no explicit task information is available and any change/shift in data distribution is identified by the system on its own. Various methods are developed to mitigate forgetting in the absence of task-specific information. These methods efficiently identify and respond to any deviation in the data distribution. It is more challenging to address forgetting in task-free continual learning.

Based on the number of passes means whether the data processed by the model only once or iterated over the data multiple times CL can be classified as online and offline. Data is processed by the model in a single pass or multiple passes. Offline continual learning requires enough computational and storage resources and it has been widely studied. Whereas in online continual learning, there is limited access to past data which it difficult to remember knowledge and reduce forgetting. The presence of limited processing capabilities or memory makes it more challenging to treat forgetting in online CL.

Based on the amount of labeled data available CL can be classified as supervised, semi-supervised, few-shot, and unsupervised CL. In Supervised, the presence of labeled data provides clear task boundaries. Difficulties arise in other forms of CL like in semi-supervised CL where only a small

amount of labeled data is available; the difficulty arises in extracting meaningful information from unlabeled data to reduce forgetting. Few-shot presents additional hurdles as there are limited labeled data points are available for learning. The lack of clear task boundaries makes unsupervised CL most challenging to identify and manage forgetting.

CL is a method where a model sequentially learns multiple tasks without forgetting prior knowledge. CL problems may be task-aware or task-free, online or offline, supervised, semi-supervised, few-shot, or unsupervised. While learning multiple tasks sequentially few challenges arise, and catastrophic forgetting is one of them.

1.3 CATASTROPHIC FORGETTING

Catastrophic forgetting (CF) is a major problem for continual learning while learning multiple tasks sequentially. It was first exposed by Mccloskey and Cohen in 1889. According to [2] Catastrophic Forgetting is the forgetting of previously learned tasks (concepts) by a neural network as it is sequentially trained on new concepts, it is responsible for degrading the performance of the machine learning model. It is also named a stability-plasticity dilemma [3]. The factors responsible for catastrophic forgetting are parameter shift, activation shift, and inter-task confusion.

Parameter shift– It is a drift in parameter to the parameter space region when the neural network is updated using gradient descent.

Activation shift –A significant change in activation of neurons when the model is trained on a new task and the performance of previous tasks degrades. Parameter shift leads to Activation shift.

Inter-task confusion -It occurs in multi-task learning. A model trained on multiple tasks is not able to differentiate between different tasks due to the interference of knowledge learned from previous tasks and new tasks. Inter-task misclassification results from the outputs and intermediate activations.

1.3.1 FACTORS OF CATASTROPHIC FORGETTING

i) **OVERFITTING:** It is an undesirable behavior where machine learning models perform well For training data but not for new data (test data). It is a challenge in which a model accurately learns the training data with its outliers and noise, but performance on new unseen data is poor. The common causes for overfitting are:

Complexity: Model complexity leads to modeling the noise in the training data.

Training Data: Small training data is not an adequate representation of the real distribution from

which it is sampled. Thus, the model is unable to learn a generalizable mapping.

ii) **INTERFERENCE:** In this process, new tasks and previous tasks are related [3]. In this, learning about the new task in the model "interferes" with the knowledge of the previous task.

iii) **ONLINE LEARNING:** In online learning, instead of training on a predetermined set of samples all at a time, the model is continuously updated with new examples as they are received. The newly received samples are different from the previous samples on which the model is trained [4]. Thus, degradation of previous task performance or forgetting of previous tasks knowledge occurs.

1.4 ADDRESSING FORGETTING

In this section, the methods of addressing forgetting based on the types of CL are discussed [4].

1.4.1 TASK-AWARE

In task-aware learning, forgetting can be addressed by five methods Memory, Architecture, Regularization, Subspace, and Bayesian methods as shown in Figure 1.

i) Memory-Based Method: In this method, various data samples from the previous tasks are stored in a buffer and these samples are replayed during learning with new tasks.

ii) Architecture-based Method: In this method, the network architecture is updated to keep the old task knowledge during the learning of new tasks [5].

iii) Regularization-based Method: This method is used to avoid forgetting old task knowledge by adding regularization loss terms to the training objective. Knowledge Distillation and Weight Regularization are the ways used to mitigate forgetting in these methods.

iv) Subspace-based Method: These methods are used to avoid forgetting that rises due to the interference among several tasks by conducting learning in independent and disjoint subspaces [6].

v) Bayesian Method: These methods provide practical ways to reduce forgetting by using regularization and uncertainty estimation techniques. The adaptability of the learning is also enhanced by these methods.

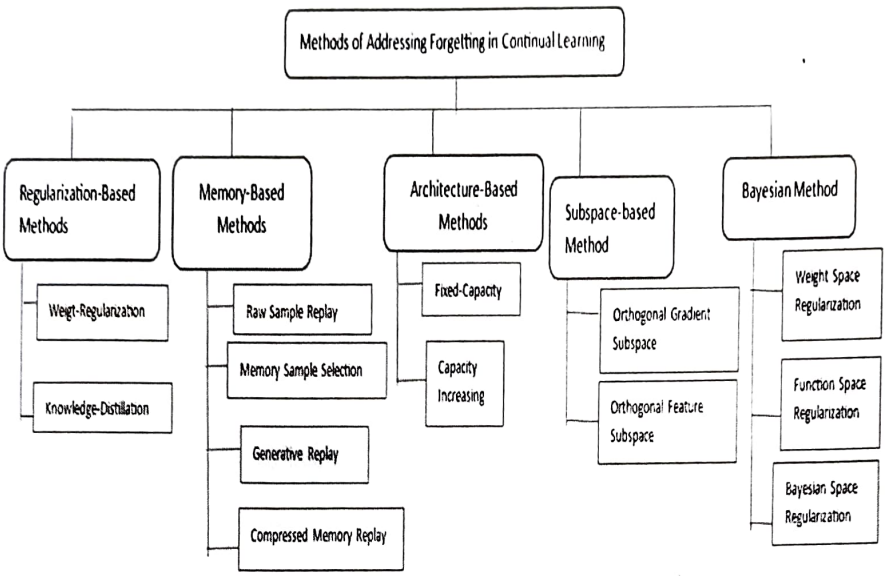

Figure 1: Methods of Reducing Forgetting [4]

1.4.2 TASK-FREE

There are two methods of addressing forgetting in task-free Continual Learning based on Memory and Network expansion as discussed below:

i) Memory-based method: In these methods, a small subset of previously stored data is replayed along with new mini-batch data.

ii) Expansion-based method: In these methods architecture of the model is expanded to avoid forgetting and enhance the ongoing adaptation.

1.4.3 ONLINE CONTINUAL LEARNING

In online CL, data for each task is processed once. It is more challenging instead of offline CL [6]. There are two methods of addressing forgetting in online Continual Learning: Rehearsal Replay and Regularization. Replay online methods are not very effective in situations where memory restrictions are severe, here regularization methods are used to prevent forgetting.

1.4.4 SEMI-SUPERVISED, FEW-SHOT AND UNSUPERVISED

The methods of addressing forgetting in semi-supervised, few-shot, and unsupervised CL for small-label training datasets are:

Semi-supervised CL: It is a variant of classical CL that permits each task to include labeled as well as unlabeled data. It includes generative replay and distillation methods to mitigate forgetting. Few-Shot CL: In Few-shot CL model new tasks are learned with limited labeled data per task while retaining the knowledge from previously learned tasks [7]. It encounters the problem of overfitting. It includes methods to mitigate forgetting such as metric-learning, meta-learning, and parameter regularization.

i) Metric Learning-Based: In these methods, Classification is performed by class prototypes in which new and old class prototypes are separated to prevent forgetting [7]. During the adjustment process of the new class, the old class prototype should not change very much.
ii) Meta-Learning-Based: These methods use inference during training, and the models adapt rapidly to new classes to solve few-shot CL.
iii) Parameter Regularization-Based: These are used to address forgetting problems like Knowledge Distillation and Penalizing parameter updates [8].

Unsupervised CL: It is a fast-growing research area that uses unlabeled data for learning and adaptation. Current unsupervised CL methods rely on contrastive learning techniques based on representation [9].

In conclusion, methods of mitigating forgetting in task-free, task-aware, supervised, unsupervised, online, and few-shot CL has been widely used by many researchers to address forgetting.

1.5 MOTIVATION

There are several factors behind the motivation of continual learning. CL has adaptability in dynamic environments, mitigation of catastrophic forgetting, resource efficiency, and scalability. Further to reduce forgetting in CL the concept of transfer learning, concept drift, and domain shift has been widely applied. Reduced forgetting enhances the performance of the system in various domains like robotics, natural language processing, finance, education, health care, autonomous vehicles, gaming and virtual reality, etc.

2. LITERATURE REVIEW

The work related to methods based on Replay, Regularization, and architecture are discussed in this section.

2.1 RESEARCH WORK ON REPLAY METHODS

In this subsection, the research on replay methods has been discussed.

Wojciech Masarczyk et al. [10] used synthetic data in sequence to reduce Catastrophic forgetting that occurs in neural networks. The authors proposed a method to generate synthetic data in a two-step optimization process via meta-gradients. The experiment carried out on the Split-MNIST dataset showed that training a model on synthetic data in sequence does not result in forgetting.

Hyuntak Cha et al. [11] found that contrastively learned representations are better at reducing forgetting than models trained by cross-entropy objectives. The authors discussed that rehearsal-based learning focuses on learned representations using contrastive learning and conserves these representations using a self-supervised distillation method.

Dushyant Rao et al. [12] introduced a Continual Unsupervised Representation Learning (CURL) approach which described the problem of the unsupervised CL where the task boundaries and label are unidentified. The proposed approach performs task inference by the use of a mixture of Gaussian latent space. To reduce catastrophic forgetting and instantiate new concepts, it uses a mixture of generative replay (MGR) and dynamic expansion. Experiments on MNIST and Omniglot proved that in a sequential class scenario, CURL exhibits no forgetting while learning meaningful class-discriminative representations.

Rishabh Tiwari et al. [13] used a CL approach based on replay to address catastrophic forgetting, which keeps a small buffer to store data samples from previously encountered tasks. The authors proposed the Gradient Corset Replay (GCR) method for replay buffer selection and update by using optimization criteria. The authors found supervised contrastive loss for continual yields a cumulative gain of up to 5% accuracy.

Gido M. van de Ven et al. [14] discussed performance, diagnostics, and resource efficiency metrics to evaluate CL approaches. The authors used six computational approaches in which replay, parameter regularization, and functional regularization modify the loss function, optimization approaches optimize the loss function, contextual and task-specific information-based computation is spread over the networks for context-based processing and template-based classification distinguished the object that has no clear boundaries. The authors also observed that deep learning handles the Challenges caused by Catastrophic forgetting and non-stationary environments.

Huong-Giang Doan et al. [15] developed an efficient solution for CF in CL by using two strategies (i) random sample selection for replay memory to avoid overfitting and (ii) end-to-end continual learning with loss combination to reduce catastrophic forgetting in DNN. The experiment showed outstanding performance and gained recognition accuracy in a range of 1% to 16%. A few limitations related to memory for sample storage and the architecture used is ResNet18. Multi-criteria strategy for memory reconstruction is used to enhance the framework.

Sajal Dash et al. [16] proposed an approach to construct a historical summary that is used with newly coming data. The authors introduced various data summarization techniques such as K-means clustering, random sampling, coreset computation, and auto encoders to reduce catastrophic forgetting. The authors found that the proposed method performs better than EWC and vanilla SGD to mitigate catastrophic forgetting. The authors also observed catastrophic forgetting is evident in ANN and CNN but not in an RNN.

2.2 RESEARCH WORK ON REGULARIZATION METHODS

In this subsection, the research on regularization methods has been discussed.

James Kirkpatrick et al. [3] developed a task-specific synaptic consolidation algorithm called elastic weight consolidation (EWC). The authors concluded that EWC is used effectively in supervised learning and reinforcement learning problems to train several tasks continuously without forgetting. The authors also demonstrated a scalable technique through the completion of a series of classification tasks using the MNIST handwritten digit dataset and sequential learning of several Atari 2600 games.

Ian J. Goodfellow et al. [6] conducted a study on training algorithms based on gradient and activation functions. The authors found that the dropout algorithm is best in all senses like retaining the old tasks and learning the new task and a fine trade-off curve exists between them. The authors also suggest the activation function should always be cross-validated.

Seyed Iman Mirzadeh et al. [8] studied how the batch size, learning rate decay, and dropout rate affect the formation of training regimes that widen the local minima of tasks which enhances the network stability and reduces forgetting.

Mohammed Amer et al. [17] demonstrate a modular neural network topology named dynamic

information balancing (DIB) across several modules by routing inputs according to the information contained in each module. The proposed technique performs better when combined with elastic weight consolidation (EWC) for a variety of lifelong learning datasets and tasks of comparable capacity.

Chenze Shao et al. [18] observed catastrophic forgetting occurs when machine translation models are trained with static training datasets. In low-resource translation forgetting causes imbalance training and affects the translation quality. The authors proposed complementary online knowledge distillation (COKD) an approach to address the imbalance training issue and obtain significant gain in translation quality.

Giang Nguyen et al. [19] introduced a technique Auto DeepVis to dissect catastrophic forgetting by knowing the plastic components of a model. The authors observed the forgetting layer and by freezing critical layers provides significant performance on both previous and coming tasks.

Prakhar Kaushik et al. [20] proposed relevance mapping to reduce the problem of forgetting and remembering by reflecting the significance of the weights for the task and allot large weights to important parameters of the neural network.

Divyam Madaan et al. [21] focused on unsupervised continual learning (UCL) and learning the feature representations on an unlabeled sequence of tasks. The authors demonstrated UCL representations that are more robust to catastrophic forgetting, perform well on few-shot learning tasks, and generalize ODD (out of distribution) tasks than supervised continual learning (SCL). The authors also proposed a lifelong unsupervised mixup an effective technique that interpolates the current and previous tasks to alleviate catastrophic forgetting for unsupervised representations.

Benjamin Ehret et al. [22] applied EWC to RNNs and found high working memory is required. It also ported the CL technique effectively to RNN and observed that regularization based on hypernetwork outperformed weight importance methods.

Pietro Buzzega et al. [23] proposed a simple baseline for CL called Dark Experience Replay that uses Knowledge Distillation combined with rehearsal and regularization for preserving past experiences to avoid catastrophic forgetting. The authors concluded that it performs better than integrated approaches and uses limited resources.

Mohammad Samin Yasar et al. [24] developed a new Continual Representation Learning framework for Overcoming Catastrophic Forgetting (CoRaL) to address forgetting. It consists of a Representation Learning module (for Plasticity) and a Knowledge Distillation module (for Stability) to provide the perfect mix of stability and plasticity in the CL Scenario. It was evaluated on three CL scenarios and four datasets and found that CoRaL is better than state-of-the-art methods.

2.3 RESEARCH WORK ON ARCHITECTURE-BASED METHODS

In this subsection, the research on architecture-based methods has been discussed.

Seyed Iman Mirzadeh et al. [5] discussed the effect of the "width" of the neural network architecture on catastrophic forgetting. It is carried out by learning the dynamics of the network from different viewpoints like gradient orthogonality, sparsity, and lazy training regime.

Xilai Li et al. [25] proposed a framework for handling CF in continual learning with DNNs. The framework consists of a neural structure optimization component and a parameter learning component. In the neural structure component each layer has three options reuse, adapt, and new in learning a new task. In the parameter learning component, fine-tuning of the reuse layer is carried out.

Sachin Gaur et al. [26] proposed the variants of the Resnet model and studied the effects of structural attributes like shape, depth, width and use of max-pooling on model performance which explore the strength and area of further improvement towards forgetting in continual learning. The authors find wider architecture with a balanced depth-to-width ratio is good for reducing forgetting. The limitations to be addressed in the future, first only one dataset and one model are used, and second fixed architecture is used.

Qiang Gao et al. [27] developed a novel technique for the Continual learning problem called Continual Learning with Efficient Architecture Search (CLEAS) which uses neural architecture search (NAS) as an efficient architecture search. NAS uses Reinforcement Learning to find the optimal architecture for each task dynamically. The technique completely alleviates catastrophic forgetting by not changing the old neurons and their optimized weights, and it uses simple architecture. This technique showed outstanding performance; its main limitation is the large running time.

Vishal Thengane et al. [28] evaluated Contrastive Language-Image Pretraining (CLIP), a simple and effective baseline model under three incremental learning scenarios task-agnostic, class-incremental, and domain-incremental with different settings. The experiments performed on five benchmark datasets (ImageNet-100 & 1K, CORe50, CIFAR-100, and TinyImageNet) showed outstanding performance against state-of-the-art CL approaches in maximum settings without any fine-tuning, replay buffers, or memory overhead. The information leakage is one of its limitations. To enhance performance in the future, CLIP will be integrated with memory replay or regularization paradigms.

2.4 RESEARCH WORK ON META-LEARNING METHOD

In this subsection, the research on meta-learning methods has been discussed.

Pauching Yap et al. [7] proposed a gradient-based model-agnostic meta-learning algorithm (MAML). In which catastrophic forgetting in classification problems is reduced by introducing a Bayesian online meta-learning framework along with Laplace approximation and variational inference.

Zhenyi Wang et al. [29] developed a meta optimizer in sequential domain meta-learning (SDML) to reduce forgetting problems, and meta optimizer is applied to both simple domain-aware meta-learning and general domain-agnostic meta-learning settings. A benchmark of 10 heterogeneous domains having 100K tasks is also proposed by them. The experiment states that the proposed method outperforms over strong baseline by a big margin.

2.5 RESEARCH WORK ON GRADIENT-BASED METHODS

In this subsection, the research on gradient-based methods has been discussed.

Zixuan Ke et al. [9] proposed a novel technique Capsules and transfer routing for continual learning (CTR) using the pre-trained model Bidirectional Encoder Representations from Transformers (BERT) in task continual learning to achieve both catastrophic forgetting avoidance and knowledge transfer. It leverages the pre-trained BERT for continual learning.

Wei Shiung Liew et al. [30] described the impact of various pruning methods on the episodic and semantic memory in Growing-Dual-Memory (GDM) networks. The performance of these methods, like synapse-aging and neuron habitation etc. are evaluated by using various performance

metrics such as neuron count, synapse count, testing accuracy of previous and current knowledge etc., and characteristics of the GDM network. The authors considered factors like age, habituation, and significance before making a pruning decision.

David Lopez-Paz et al. [31] proposed an approach to reduce forgetting by transferring knowledge forward and backward between tasks. The authors developed a model based on Gradient Episodic Memory (GEM) using MNIST and CIFAR-100 datasets.

Haiyan Zhao et al. [32] investigated parts of a neural network that are task-specific and more prone to catastrophic forgetting. The authors proposed a simple and effective method "forgetting prioritized fine-tuning (FPF)" which fine-tunes a subset of parameters on a small buffer before the deployment of the model. To improve efficiency, they replaced replay with k-FPF and attained comparable performance by taking approximately half of its computation.

Irene Munoz-Martin et al. [33] described the spike-timing-dependent plasticity (STDP) concept and neural redundancy to reduce catastrophic forgetting. STDP is explained by introducing one layer perceptron adopting phase change memory (PCM) synapses. A hybrid neural network architecture is proposed by authors that blends supervised and unsupervised learning to tackle the "stability/plasticity dilemma" related to catastrophic forgetting in neural networks. Different tasks and datasets impact network performance, and dependence on specific memory devices, such as phase-change memory, limits the network's generalizability and scalability.

Arslan Chaudhry et al. [34] introduced a more realistic evaluation protocol that selects the hyper-parameters on a small and distinct collection of tasks and new metric measuring for quickly acquiring a new skill. The authors proposed A-GEM improved version of GEM performed better as computationally and memory efficient as EWC. The authors found that by providing the task descriptors A-GEM learn quickly and perform better than any other method. Task descriptors enable both the learning curve area (LCA) and the few-shot learning regime.

In conclusion, depending upon the problems faced by researchers' various techniques are used to overcome forgetting in continual learning. The models are trained by using regularization, replay, parameter isolation, and gradient-based methods. In some situations, a hybrid approach is used to reduce forgetting.

3. RESEARCH GAPS

Although many researchers have done work in reducing catastrophic forgetting still there is a need for further improvement to make the continual learning model adaptable and robust. Some research gaps have been identified as given below:

1. The large volume of data, complexity, and forgetting should be focused on while training the model.
2. Proper trade-off between Memorization and Forgetting is required.
3. Limited theoretical frameworks are available to explain the mechanism behind catastrophic forgetting and continual learning.
4. Lack of strategies to optimally balance stability and plasticity in a dynamic learning environment.
5. The efficiency of existing algorithms needs to be improved.
6. There is a requirement to create a task-agnostic algorithm for reducing forgetting.
7. Require improvement in efficient memory usage and memory augmentation.

By addressing these research gaps, models become more efficient and robust, and the system learns continuously without forgetting.

4. PROBLEM STATEMENT

Continual learning in AI enables a model to learn from new tasks without forgetting previous tasks. A major challenge is catastrophic forgetting, where new learning overwrites old knowledge and degrades the performance of the model on previous tasks. Existing methods such as regularization, replay, and modular architectures address this problem, but none has resolved it completely. This research designs a novel machine learning technique for reducing catastrophic forgetting in continual learning for improving knowledge retention. An empirical evaluation of the proposed solution would be conducted using benchmark datasets to determine its efficacy and practicality.

5. RESEARCH OBJECTIVES

The objectives of the research are as given below:

1. To study the literature related to various methods of continual learning.
2. To analyze the methods based on supervised learning for identifying the level of forgetting.

3. To develop a technique for reducing the level of forgetting in continual learning.
4. To empirically evaluate the developed technique based on a benchmark dataset.

6. RESEARCH METHODOLOGY

The following methodology will be used to carry out the research: -

1. A literature review of various continual learning methods and their comparison will be performed.
2. Preprocessing of the benchmark dataset will be done using deep learning techniques.
3. Propose a technique to reduce forgetting in continual learning.
4. Compare the developed technique with existing state of the art methods and identify the level of reduction in forgetting.
5. The developed technique will be evaluated on a benchmark dataset.

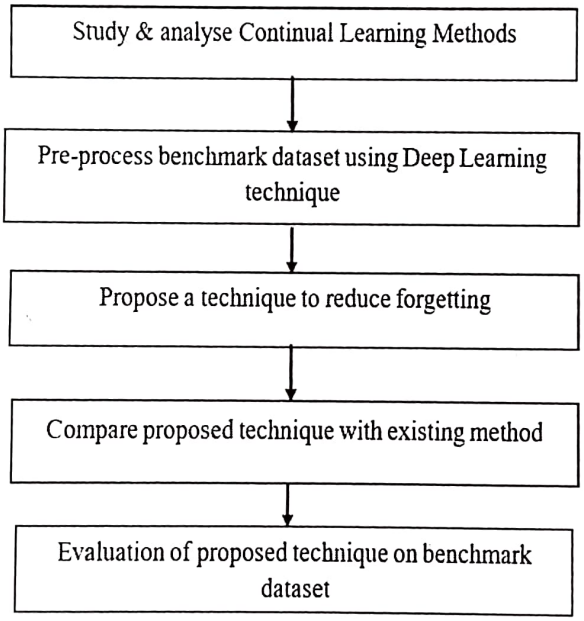
Study & analyse Continual Learning Methods
Pre-process benchmark dataset using Deep Learning technique
Propose a technique to reduce forgetting
Compare proposed technique with existing method
Evaluation of proposed technique on benchmark dataset

Figure 2: Research Methodology

7. EXPECTED OUTCOME OF RESEARCH

This research aims to give a technique using machine learning to reduce catastrophic forgetting in continual learning to enhance the performance of the model.

REFERENCES:

[1] I. H. Sarker," Machine Learning: Algorithms, Real-World Applications and Research Directio -ns," SN Computer Science, vol. 2, no. 3, pp. 1–21, 2021, doi: https://doi.org/10.1007/s42979-021-00592-x.

[2] Z. Chen and B. Liu," Continual Learning and Catastrophic Forgetting," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, pp. 55–75, 2018, doi: https://doi.org/10.1007/978-3-031-01581-6_4.

[3] J. Kirkpatrick et al.,"Overcoming catastrophic forgetting in neural networks", Proceedings of the National Academy of Sciences, vol. 114, no. 13, pp. 3521–3526. Mar. 2017. doi:https://doi.org/10.1073/pnas.1611835114

[4] Z. Wang, E. Yang, L. Shen, and H. Huang, "A Comprehensive Survey of Forgetting in Deep Learning Beyond Continual Learning," *arXiv.org*, Jul. 23, 2023. https://arxiv.org/abs/2307.09218

[5] S. I. Mirzadeh et al., "Wide Neural Networks Forget Less Catastrophically," Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA. PMLR,162, Jun., 28, 2022. https://proceedings.mlr.press/v162/mirzadeh22a.

[6] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks," arXiv:1312.6211 [cs, stat], Mar. 2015, doi: https://doi.org/10.48550/arXiv.1312.6211.

[7] P. Yap, H. Ritter, and D. Barber, "Addressing Catastrophic Forgetting in Few-Shot Problems," arXiv.org, Jun. 21, 2021. https://arxiv.org/abs/2005.00146 (accessed Jun. 02, 2024).

[8] S. I. Mirzadeh, M. Farajtabar, R. Pascanu, and H. Ghasemzadeh, "Understanding the role of training regimes in continual learning: 34th Conference on Neural Information Processing Systems, NeurIPS 2020," Advances in Neural Information Processing Systems, vol. 2020 December, 2020. Available: https://proceedings.neurips.cc/paper/2020/file/518a3 8cc9a0173 d0b2dc088166981cf8-Paper.pdf.

[9] Z. Ke, B. Liu, N. Ma, H. Xu, and L. Shu, "Achieving Forgetting Prevention and Knowledge Transfer in Continual Learning," 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Dec. 2021, doi: https://doi.org/10.48550/arxiv.2112 0270.

[10] W. Masarczyk and I. Tautkute, "Reducing catastrophic forgetting with learning on synthetic data," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 1019-1024, doi: 10.1109/CVPRW50498.2020.00134.

[11] H. Cha, J. Lee, and J. Shin, "Co2L: Contrastive Continual Learning," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 9496-9505, doi:10.1109/ICCV48922.2021.00938.

[12] D. Rao, F. Visin, A. A. Rusu, Razvan Pascanu, Yee Whye Teh, and R. Hadsell, "Continual Unsupervised Representation Learning.," *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.*, vol. 32, pp. 7645–7655, Jan. 2019.

[13] R. Tiwari, K. Killamsetty, R. K. Iyer, and P. Shenoy, "GCR: Gradient Coreset based Replay Buffer Selection for Continual Learning," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, doi: https://doi.org/10.1109/cvpr52688.2022.00020.

[14] G. M. van de Ven, N. Soures, and D. Kudithipudi, "Continual Learning and Catastrophic Forgetting," *arXiv.org*, Mar. 08, 2024. https://arxiv.org/abs/2403.05175.

[15] H.-G. Doan, H.-Q. Luong, T.-O. Ha, and T. Thanh, "An Efficient Strategy for Catastrophic Forgetting Reduction in Incremental Learning," *Electronics*, vol. 12, no. 10, pp. 2265–2265, May 2023, doi: https://doi.org/10.3390/electronics12102265.

[16] S. Dash, J. Yin, M. Shankar, F. Wang, and W. Feng, "Mitigating Catastrophic Forgetting in Deep Learning in a Streaming Setting Using Historical Summary," OSTI OAI (U.S. Department of Energy Office of Scientific and Technical Information), 7th international workshop on data analysis and reduction for Big Scientific Data (2021)., Nov. 2021, doi: https://doi.org/10.1109/drbsd754563.2021.00006.

[17] M. Amer and T. Maul, "Reducing Catastrophic Forgetting in Modular Neural Networks by Dynamic Information Balancing," arXiv.org, Dec. 10, 2019. https://arxiv.org/abs/1912.04508.

[18]C. Shao and Y. Feng, "Overcoming Catastrophic Forgetting beyond Continual Learning: Balanced Training for Neural Machine Translation," ACL Web, May 01, 2022. https://aclanthology.org/2022.acl-long.143/.

[19]G. Nguyen, S. Chen, T. Do, T. J. Jun, H.-J. Choi, and D. Kim, "Dissecting Catastrophic Forgetting in Continual Learning by Deep Visualization," arXiv.org, Jan. 07, 2020. https://arxiv.org/abs/2001.01578.

[20] P. Kaushik, A. Gain, A. Kortylewski, and A. Yuille, "Understanding Catastrophic Forgetting and Remembering in Continual Learning with Optimal Relevance Mapping," arXiv:2102.11343 [cs], Feb. 2021, Available: https://arxiv.org/abs/2102.11343.

[21] D. Madaan, J. Yoon, Y. Li, Y. Liu, and S. J. Hwang, "Representational Continuity for Unsupervised Continual Learning," arXiv (Cornell University), Oct. 2021, doi: https://doi.org/10.48550/arxiv.2110.06976.

[22] B. Ehret, C. Henning, M. Cervera, A. Meulemans, J. Von Oswald, and B. Grewe, "Published as a conference paper at ICLR 2021 CONTINUAL LEARNING IN RECURRENT NEURAL NETWORKS," Jan. 2021. Accessed: Jun. 02, 2024. [Online]. Available: https://openreview.net/pdf?id=8xeBUgD8u9.

[23] P. Buzzega, M. Boschini, A. Porrello, D. Abati, and S. Calderara, "Dark Experience for General Continual Learning: a Strong, Simple Baseline," *arXiv (Cornell University) 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada*, Apr. 2020, doi: https://doi.org/10.48550/arxiv.2004.07211.

[24] M. Samin Yasar and T. Iqbal, "CoRaL: Continual Representation Learning for Overcoming Catastrophic Forgetting," International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, May 2023. Available: https://www.ifaamas.org/Proceedings/aamas2023/pdfs/p1969.

[25] X. Li, Y. Zhou, T. Wu, R. Socher, and C. Xiong, "Learn to Grow: A Continual Structure Learning Framework for Overcoming Catastrophic Forgetting," *arXiv.org*, May 21, 2019. https://arxiv.org/abs/1904.00310.

[26] S. Gaur, "Efficient and Effective Architecture in Continual Learning Through Various ResNets," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 1148–1161, Mar. 2023, Available: https://ijisae.org/index.php/IJISAE/article/view/5394.

[27] Q. Gao, Z. Luo, D. Klabjan, and F. Zhang, "Efficient Architecture Search for Continual Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 1–11, 2022, doi: https://doi.org/10.1109/tnnls.2022.3151511.

[28] Vishal Thengane, S. Khan, M. Hayat, and F. Khan, "CLIP model is an Efficient Continual Learner," *OpenReview*, 2022. https://openreview.net/forum?id=qZ1yqKvQtE.

[29] Z. Wang, L. Shen, T. Duan, D. Zhan, L. Fang, and M. Gao, "Learning to Learn and Remember Super Long Multi-Domain Task Sequence," in CVPR, 2022. Available: https://openaccess.thecvf.com/content/CVPR2022/papers/Wang_Learning_To_Learn_and_Remember_Super_Long_Multi-Domain_Task_Sequence_CVPR_2022_paper.pdf

[30] W. S. Liew, C. Kiong Loo, V. Gryshchuk, C. Weber and S. Wermter, "Effect of Pruning on Catastrophic Forgetting in Growing Dual Memory Networks," *2019 International Joint*

Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8851865.

[31] D. Lopez-Paz and Marc'Aurelio Ranzato, "Gradient Episodic Memory for ContinualLearning,"31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA, Jun. 2017, doi: https://doi.org/10.48550/arxiv.1706.08840.

[32] H. Zhao, T. Zhou, G. Long, J. Jiang, and C. Zhang, "Does Continual Learning All Parameters?" Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023., Apr. 2023, doi: https://doi.org/10.48550/arxiv.2304.04158.

[33] I. Munoz-Martin, S. Bianchi, G. Pedretti, O. Melnic, S. Ambrogio, and D. Ielmini, "Unsupervised Learning to Overcome Catastrophic Forgetting in Neural Networks," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 5, no. 1, pp. 58–66, Jun. 2019, doi: https://doi.org/10.1109/jxcdc.2019.2911135.

[34] A. Chaudhry, M. Ranzato, M. Rohrbach, and M. Elhoseiny, "Efficient Lifelong Learning with A-GEM," *arXiv.org*, Jan. 09, 2019. https://arxiv.org/abs/1812.00420 (accessed Aug. 27, 2023).

MANAV RACHNA
vidyanatariksha

Dr. O.P. Bhalla Central Library
Manav Rachna International Institute of Research and Studies
(A deemed-to-be-university under section 3 of UGC Act 1956)

SIMILARITY VERIFICATION CERTIFICATE

Title: Designing Machine learning Technique For Reducing Catastrophic Forgetting in Continual Learning

Researcher: Neeraj

Registration Number: 22-23/Ph.D/W/018

Supervisor: Dr. Poonam Nandal

Department: Department of Computer Science and Engineering

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Turnitin Software on 11/11/25 with submission ID 1:3406385563 and the similarity index was found to be 6 %. The complete report running into 23 Pages, 6489 Word count with File Size 210.2 KB/MB is being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 11/11/25

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International Institute of Research and Studies
Faridabad

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Research Scholar

Research Supervisor

11/11/25

Neeraj .

DESIGNING MACHINE LEARNING TECHNIQUE FOR REDUCING CATASTROPHIC FORGETTING IN CONTINUAL LE...

Quick Submit

Quick Submit

Manav Rachna International University

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

Document Details

Submission ID
trn:oid:::1:3406385562

Submission Date
Nov 11, 2025, 10:48 AM GMT+5:30

Download Date
Nov 11, 2025, 10:58 AM GMT+5:30

File Name
FOR_REDUCING_CATASTROPHIC_FORGETTING_IN_CONTINUAL_LEARNING.docx

File Size
210.2 KB

23 Pages

6,489 Words

39,796 Characters

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Small Matches (less than 12 words)

Match Groups

- **17** Not Cited or Quoted **4%**
  Matches with neither in-text citation nor quotation marks
- **4** Missing Quotations **1%**
  Matches that are still very similar to source material
- **1** Missing Citation **0%**
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted **0%**
  Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 3% Publications
- 2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bindra Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

Match Groups

- 17 Not Cited or Quoted 4%
  Matches with neither in-text citation nor quotation marks
- 4 Missing Quotations 1%
  Matches that are still very similar to source material
- 1 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 3% Publications
- 2% Submitted works (Student Papers)

11/11/25

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
Bahcesehir University		<1%
2	Internet	
www.osti.gov		<1%
3	Student papers	
Manav Rachna International University		<1%
4	Publication	
Li, Xilai. "Neural Architectures Design and Search for Deep Learning and Continu...		<1%
5	Student papers	
University of Nottingham		<1%
6	Publication	
Wei Shiung Liew, Chu Kiong Loo, Vadym Gryshchuk, Cornelius Weber, Stefan Wer...		<1%
7	Internet	
bridgingthegaps.bris.ac.uk		<1%
8	Internet	
www.mdpi.com		<1%
9	Internet	
dynn-icml2022.github.io		<1%
10	Internet	
openaccess.thecvf.com		<1%

11	Student papers	
Chaffey College		<1%
12	Student papers	
University of New South Wales		<1%
13	Internet	
arxiv.org		<1%
14	Internet	
github.com		<1%
15	Publication	
Singh, Vaibhav. "Wake-Sleep Energy Based Models for Continual Learning", New ...		<1%
16	Student papers	
University of North Texas		<1%
17	Publication	
Xiuwei Chen, Xiaobin Chang. "Dynamic Residual Classifier for Class Incremental L...		<1%
18	Publication	
Yanbei Chen, Massimiliano Mancini, Xiatian Zhu, Zeynep Akata. "Semi-Supervised...		<1%
19	Internet	
export.arxiv.org		<1%
20	Internet	
www.geeksforgeeks.org		<1%

DESIGNING MACHINE LEARNING TECHNIQUE FOR REDUCING CATASTROPHIC FORGETTING IN CONTINUAL LEARNING

SHORT SYNOPSIS

by

Neeraj

Registration number: 22-23/Ph.D/W/018

Under the supervision of

Dr. Poonam Nandal

Professor

Department of Computer Science and Engineering

School of Engineering & Technology

Manav Rachna International Institute of Research & Studies

Sector-43, Surajkund–Delhi Road, Faridabad– 121004

Abstract

Artificial intelligence is a field of computer science that depends on the ability of a machine to learn tasks sequentially, which is called continual learning. Continual learning, also known as incremental learning, where the systems are supposed to learn new knowledge progressively over time without forgetting earlier learned knowledge. One of the main obstacle to continual learning is catastrophic forgetting, in which the neural network model, when learning various tasks sequentially, loses its previously learned knowledge from dynamic data distributions and degrades the performance of earlier learned tasks. This problem has widely occurred in the connectionist model. The neural network training and stability-plasticity dilemma are affected by catastrophic forgetting, where the stability-plasticity dilemma states that when the network becomes too plastic, it lacks the stability required to prevent the previous knowledge from being forgotten. Further, it is vital to reduce catastrophic forgetting in environments where the model must perform in dynamic, real-world situations, guaranteeing that machine learning systems maintain consistent, reliable performance across multiple tasks. There are several

turnitin Page 6 of 27 - Integrity Submission Submission ID trn:oid:::1:3406385562

existing methods such as replay, regularization, and modular network architecture, to reduce catastrophic forgetting. Still, none of them has been able to successfully and completely reduce the issue in all tasks and domains. This research aims to examine the existing methods of reducing catastrophic forgetting, as well as their strengths and weaknesses, and propose a hybrid technique that enhances the retention of information while learning new tasks. The proposed hybrid technique would integrate existing methods in a balanced way to address their weaknesses effectively and offer flexibility. The proposed technique would use a hybrid model, evaluated on benchmark datasets, and further will be compared with other state of the art techniques.

Keywords: Artificial Intelligence, Machine Learning, Catastrophic Forgetting, Continual Learning, Deep Neural Network.

turnitin Page 6 of 27 - Integrity Submission Submission ID trn:oid:::1:3406385562